Homework 1

Chun-Hui,Wu

October 21, 2017

Problem 1

Proof Convexity inequality by induction.

Proof. If $f$is convex on closed interval $[a,b]$, then given $\lambda \in [0,1]$,f satisfies

$$f(\lambda a + (1-\lambda b)) \leq \lambda f(a) + (1-\lambda)f(b)$$

We now prove , given $\lambda_1, \lambda_2 ... \lambda_n$ satisfies $\sum_{i=1}^{M} \lambda_i = 1, \lambda \geq 0$, f satisfies

$$f(\sum_{i=1}^{M} \lambda_i x_i) \leq \sum_{i=1}^{M} \lambda_i f(x_i)$$

The inequality hold for $M = 1, 2$.Suppose the inequality holds for $M = n$. We now check if the inequality hold for $M = n + 1$

$$\begin{aligned} f(\sum_{i=1}^{n+1} \lambda_i x_i) &= f(\sum_{i=1}^{n} \lambda_i x_i + \lambda_{n+1} x_{n+1}) && (1) \\ &= f((1-\lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} x_i + \lambda_{n+1} x_{n+1}) && (2) \\ &\leq (1-\lambda_{n+1}) f(\sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} x_i) + \lambda_{n+1} f(x_{n+1}) && (3) \\ &\leq (1-\lambda_{n+1}) \sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_n} f(x_i) + \lambda_{n+1} f(x_{n+1}) && (4) \\ &= \sum_{i=1}^{n} \lambda_i f(x_i) + \lambda_{n+1} f(x_{n+1}) && (5) \\ &= \sum_{i=1}^{n+1} \lambda_i f(x_i) && (6) \end{aligned}$$

In (3) we use the fact that $\sum_{i=1}^{n} \frac{\lambda_i}{1-\lambda_{n+1}} = 1$ since the inequality holds.
In (4) we use the assumpsion of "M=n" , since the inequality holds. Finally , we prove the convexity inequality by induction.

□

Problem 2

Derive the entropy of the univariate Gaussian.

Proof. Given Random variable X , the entropy is defined by

$$H[X] := E[I[X]] = E[-ln(X)]$$

We now consider univariate Gaussian random variable and derive its entropy. By the definition of entropy

$$H[X] = E[-ln(X)] = -\int_{\Omega} p(x)ln(p(x))dx$$

since X is a Gaussian Random Variable,

$$\begin{aligned}
H[X] &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} ln(\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}})dx && (7)\\
&= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} (ln(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{(x-\mu)^2}{2\sigma^2})dx && (8)\\
&= \frac{1}{2}ln(2\pi\sigma^2) + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \frac{(x-\mu)^2}{2\sigma^2} dx && (9)\\
&= \frac{1}{2}ln(2\pi\sigma^2) + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-a^2}{2}} \frac{a^2}{2} da && (10)\\
&= \frac{1}{2}ln(2\pi\sigma^2) + \frac{1}{2} && (11)
\end{aligned}$$

From (9) to (10) we simply let $a = \frac{x-\mu}{\sigma}$,and apply the technics of change of variable. From (10) to (11) we use the fact that $E[X^2] = \sigma^2 + \mu^2$. □

Problem 3

Evaluate the KL divergence between two Gaussian.$p(x) = N(x|\mu, \sigma^2)$ and $q(x) = N(x|m, s^2)$.

Proof. K-L divergence is defined by

$$KL(p||q) := -\int p(x)ln\frac{q(x)}{p(x)}dx = -\int p(x)ln(q(x))dx + \int p(x)ln(p(x))dx$$

Given $p(x) = N(x|\mu, \sigma^2)$ and $q(x) = N(x|m, s^2)$, we derive the K-L divergence of p and q.

$$\begin{aligned}
KL(p||q) &= -\int p(x)ln(q(x))dx + \int p(x)ln(p(x))dx && (12)\\
&= -\int p(x)ln(q(x))dx - \frac{1}{2}ln(2\pi\sigma^2) - \frac{1}{2} && (13)
\end{aligned}$$

We calculate $\int p(x)ln(q(x))dx$ below.

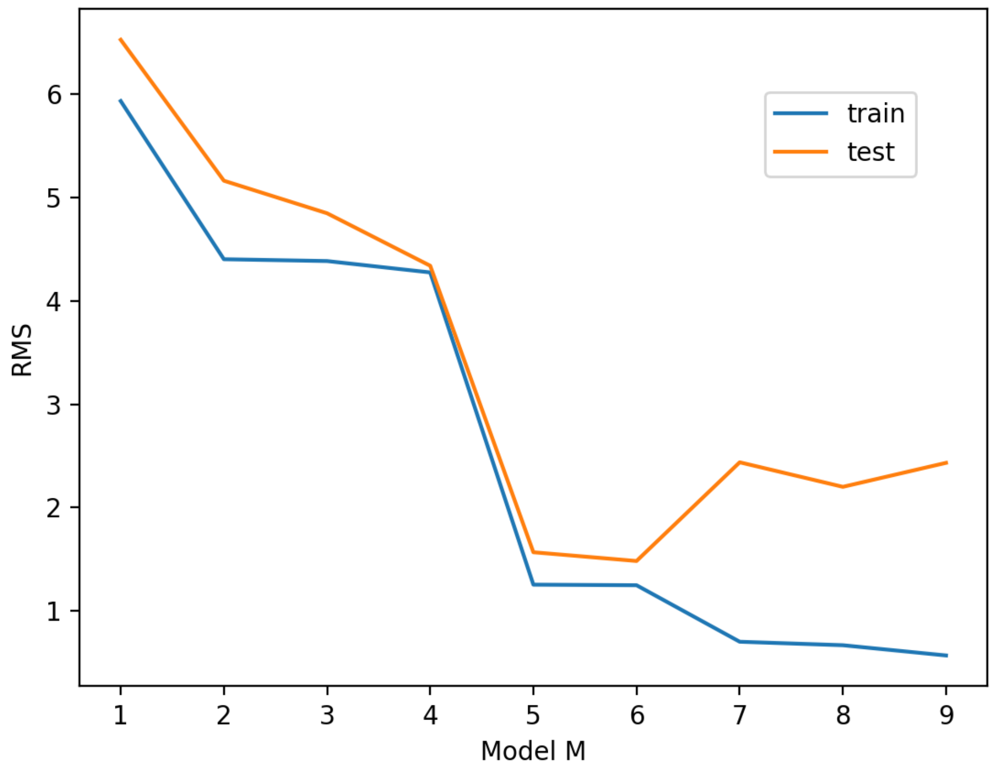
$$\begin{aligned}
\int p(x)ln(q(x))dx &= \int \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} (ln(\frac{1}{\sqrt{2\pi}s}) - \frac{(x-m)^2}{2s^2})dx && (14)\\
&= ln(\frac{1}{\sqrt{2\pi}s}) - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-a^2}{2}} \frac{(a\sigma+\mu-m)^2}{2s^2} da && (15)\\
&= ln(\frac{1}{\sqrt{2\pi}s}) - \frac{1}{2s^2}\int (\sigma^2a^2 + 2\sigma(\mu-m)a + (\mu-m)^2)\frac{1}{\sqrt{2\pi}} e^{\frac{-a^2}{2}} da && (16)\\
&= -\frac{1}{2}ln(2\pi s^2) - \frac{1}{2s^2}(\sigma^2 + (\mu-m)^2) && (17)
\end{aligned}$$

In (13) ,we use the result calculated in problem2.
In (15) ,we simply let $a = \frac{x-\mu}{\sigma}$,and apply the technics of change of variable.
Combine the result we just derive ,we know that

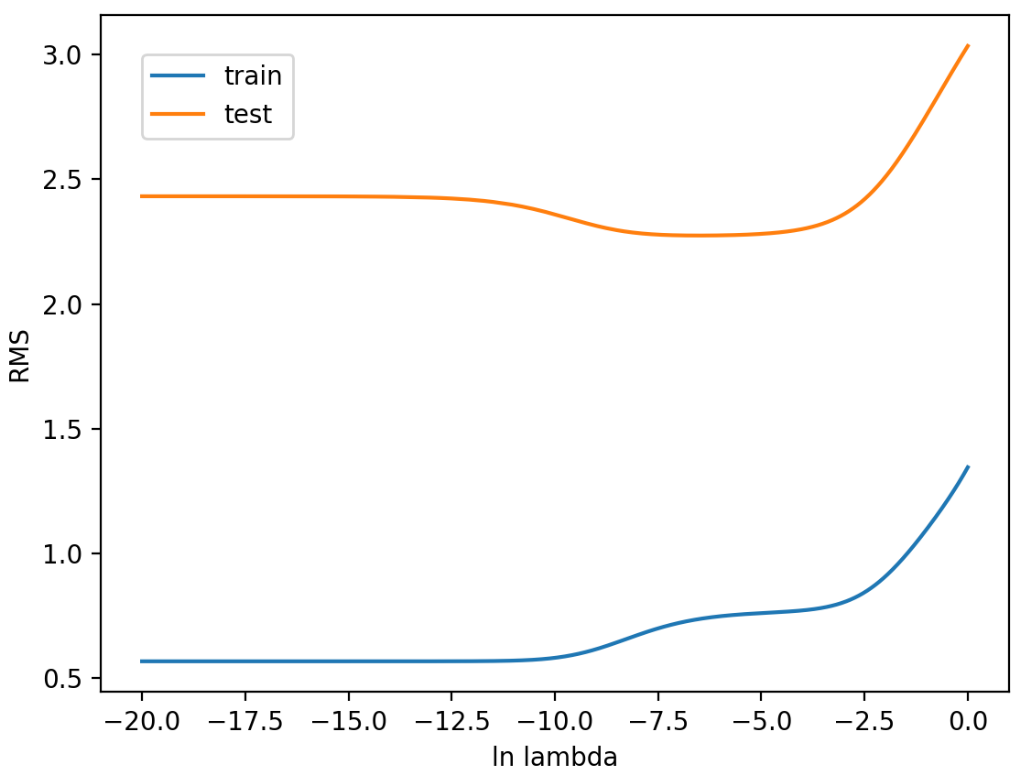
$$\begin{aligned}
KL(p||q) &= -\int p(x)ln(q(x))dx + \int p(x)ln(p(x))dx && (18)\\
&= -(-\frac{1}{2}ln(2\pi s^2) - \frac{1}{2s^2}(\sigma^2 + (\mu-m)^2)) - \frac{1}{2}ln(2\pi\sigma^2) - \frac{1}{2} && (19)\\
&= ln(\frac{s}{\sigma}) + \frac{1}{2s^2}(\sigma^2 + (\mu-m)^2) - \frac{1}{2} && (20)
\end{aligned}$$

□

Problem 4



When the order $M > 6$, overfitting occurs.



I choose 1000 $ln(\lambda_i)$ uniformly in $[-20, 0]$.

Problem 5

1

For $M = 1$,the training $RMS = 0.0512$, the testing $RMS = 0.0292$
For $M = 2$,the training $RMS = 0.0355$, the testing $RMS = 0.0349$

2

Leave petal width out the training, the training $RMS = 0.0577$, the testing $RMS = 0.0596$
Leave sepal width out the training, the training $RMS = 0.0554$, the testing $RMS = 0.0263$
Leave petal length out the training, the training $RMS = 0.0467$, the testing $RMS = 0.0269$
Leave sepal length out the training, the training $RMS = 0.0434$, the testing $RMS = 0.0305$
So if the petal width is left ,the training RMS would mostly increase among the four attributes.So petal width is the most contributive attribute.